

# SEMANTIC SEGMENTATION OF CONVOLUTIONAL NEURAL NETWORK FOR SUPERVISED CLASSIFICATION OF MULTISPECTRAL REMOTE SENSING

Xue Lei<sup>1</sup>, Liu Chang<sup>1</sup>, Wu Yunqiang<sup>1</sup>, Li Huiying<sup>1,2\*</sup>

<sup>1</sup> Computer science and technology department, Jilin University, 130012 ChangChun, Jilin province, China

<sup>2</sup> Symbol Computation and Knowledge Engineer of Ministry of Education, Jilin University, 130012 Changchun, Jilin province, China

**KEY WORDS:** Semantic Segmentation, Multi Spectral Remote Sensing, Convolutional Neural Network, U-net, multi-scale image

## ABSTRACT:

Semantic segmentation is a fundamental research in remote sensing image processing. Because of the complex maritime environment, the classification of roads, vegetation, buildings and water from remote Sensing Imagery is a challenging task. Although the neural network has achieved excellent performance in semantic segmentation in the last years, there are a few of works using CNN for ground object segmentation and the results could be further improved. This paper used convolution neural network named U-Net, its structure has a contracting path and an expansive path to get high resolution output. In the network, We added BN layers, which is more conducive to the reverse pass. Moreover, after upsampling convolution, we add dropout layers to prevent overfitting. They are promoted to get more precise segmentation results. To verify this network architecture, we used a Kaggle dataset. Experimental results show that U-Net achieved good performance compared with other architectures, especially in high-resolution remote sensing imagery.

## 1.INTRODUCTION

In recent years, some classifiers have a good performance in the image classification, such as minimum-distance classifier, Support Vector Machine (SVM), PCA linear dimension reduction method and Mean clustering method. However, their success was limited in feature extraction and classification of multispectral images.

In the last two years, deep convolutional networks have outperformed the state of the art in many visual recognition tasks. According to the characteristics of multispectral images, the convolutional network model with multi-layer perceptrons can be designed to solve this classification problem by using the spectral information and spatial information of the data.

In this paper, we present a classification method of multi-spectral images with improved U-net network. U-Net is an encoder-decoder structure in which the encoder gradually

reduces the spatial dimension of the pooling layer and the decoder gradually fixes the details and spatial dimensions of the object. There is usually a quick connection between the encoder and the decoder, so it helps the decoder to better fix the details of the target, especially in high-resolution remote sensing imagery. We classify 10 kinds of ground objects appearing in remote sensing images, and use some methods of graphics to post-process the preliminary classification results to make the classification results more accurate.

## 2.METHOD

### 2.1 Improved Network Architecture

By constructing a U-net based convolutional network model, the end-to-end correspondence of low-resolution features can be effectively achieved. U-net network combines layers of the feature hierarchy and refines the spatial precision of the output.

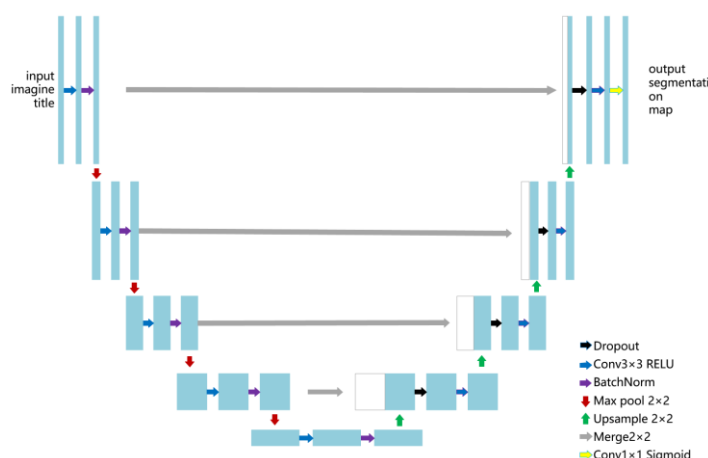


Figure 1. Our U-net architecture

\* Corresponding author Huiying Li, Email: lihuiying@jlu.edu.cn

The U-net network structure consists of a contracting path and an expansive path. The contracting path follows the typical architecture of a convolutional network. In total the network has 23 convolutional layers. It consists of the repeated application of two 3x3 convolutions (unpadded convolutions), each followed by a rectified linear unit (ReLU) and a 2x2 max pooling operation with stride 2 for downsampling. After each convoluted layer, we add a BN layer. Every step in the expansive path consists of an upsampling of the feature map followed by a 2x2 convolution (up-conv) that halves the number of feature channels, a concatenation with the correspondingly cropped feature map from the contracting path, and two 3x3 convolutions, each followed by a ReLU. After upsampling convolution, we add a dropout layer to prevent overfitting. At the final layer a 1x1 convolution is used to map each components feature vector to the desired number of classes (shown in Figure 1).

## 2.2 Data preprocessing

We have marked 10 basic objects, including housing, track, tree, etc. (Figure 2 and Table1)

a. Firstly, we will transform the coordinate, geographical coordinates of corresponding to the pixel coordinates of the picture one by one. The coordinate information is connected into an irregular polygon area, and the interior of the polygon is filled to generate a binary mask image. Housing and road renderings are as follows (Figure 3).

b. And then we combine the multi-scale image block with the sliding window generation method. The block overlap technique is used to cover the edge of the whole image. The original image and its corresponding binary mask are used as the input of the network.

c. For seven kinds of ground objects, such as houses, artificial buildings, railways, trees, crops, roads and waterlogged areas, the training set is a multi-spectral band image with 16 channels, and the size of the image is set as 8. With the goal of minimizing the sum of binary cross-entropy losses, train each category accordingly, simply average the output of all models, and then post-process the parameters according to a particular category.

d. For rivers and watersheds, a combination of linear regression and random forest can be used to identify the river and train the 8-channel input data, which works well due to the unique spectrum of water. Later post-processing by combining the indifferent moisture index and the apical chlorophyll content index gives accurate results.

e. For large and small vehicles, due to the large resolution of the training images, the amount of data relative to the vehicle is very small, we must use special means for training. Using the corresponding RGB band image data, averaging, training a fusion network, the vehicle separate segmentation.

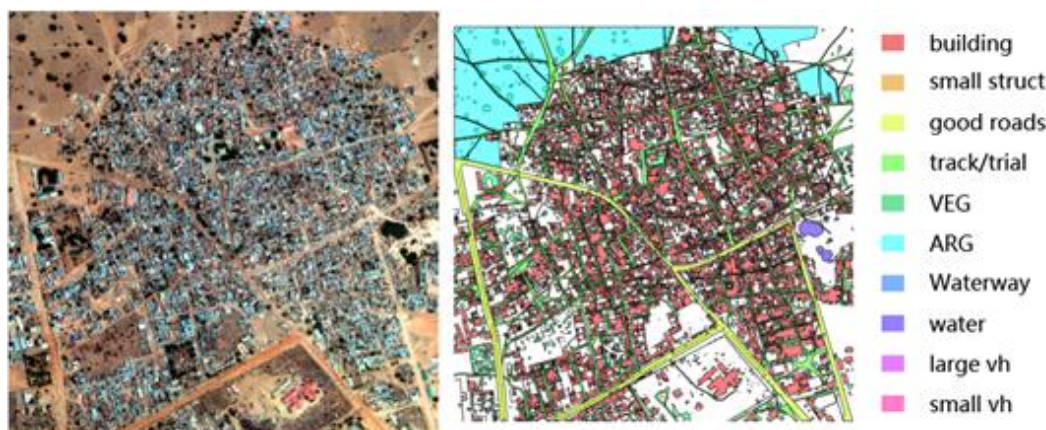


Fig.2 Segmentation result with manual ground truth

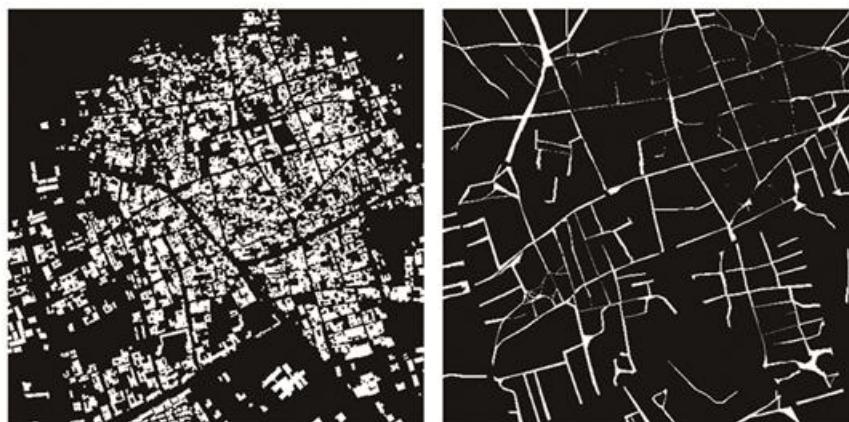


Fig.3. Preprocessing of input data.

Type	Band	Pixel resolution	Channels	Size
3 channels	RGB	0.31 m	3	3348×3392
grayscale	Panchromatic Band	0.31 m	1	3348×3392
16 channels	Multispectral	1.24 m	8	837×848
16 channels	Near-infrared band	7.5 m	8	134×136

Table 1. Parameters of image.

### 3.EXPERIMENTAL RESULTS

As shown in figure 4, a preliminary identification result of all categories of an image:

The first line displays the second to sixth channel images of original image. Line 2 and line 3 are the recognitions of each category, with red dots for the area identified, among the third figure of the second row and the Fig 3, 4, 5 of the third row show no results to identify the target figure:

As shown in Fig.5, the picture on the left is original, and the picture on the right shows the final result of the initial recognition result after the post-processing ,in which the target marked by white area.

The result of house type identification: by contrasting the effect of the left and right pictures, it can be clearly seen that each block is identified in the right picture, and even the object partially shielded by trees is effectively identified(shown in Figure 5.a).

The result of river type identification: Through the tributaries of the river clearly shown in the figure, we can see the river identification effect is obvious(shown in Figure 5.b).

The result of crop area type identification: part of the trunk road in the crop area is clearly divided, indicating that the identification of the crop area is clear(shown in Figure 5.c).

The result of main roads type identification: from the complex scene, we can clearly see the main road. We made a good identification of the main road(shown in Figure 5.d).

The result of track type identification: for the identification of subtle roads is lack. We will do more in the follow-up promotion(shown in Figure 5.e).

The result of lake type identification: Through the tributaries of the lake clearly shown in the figure, we can see the river identification effect is obvious(shown in Figure 5.f).

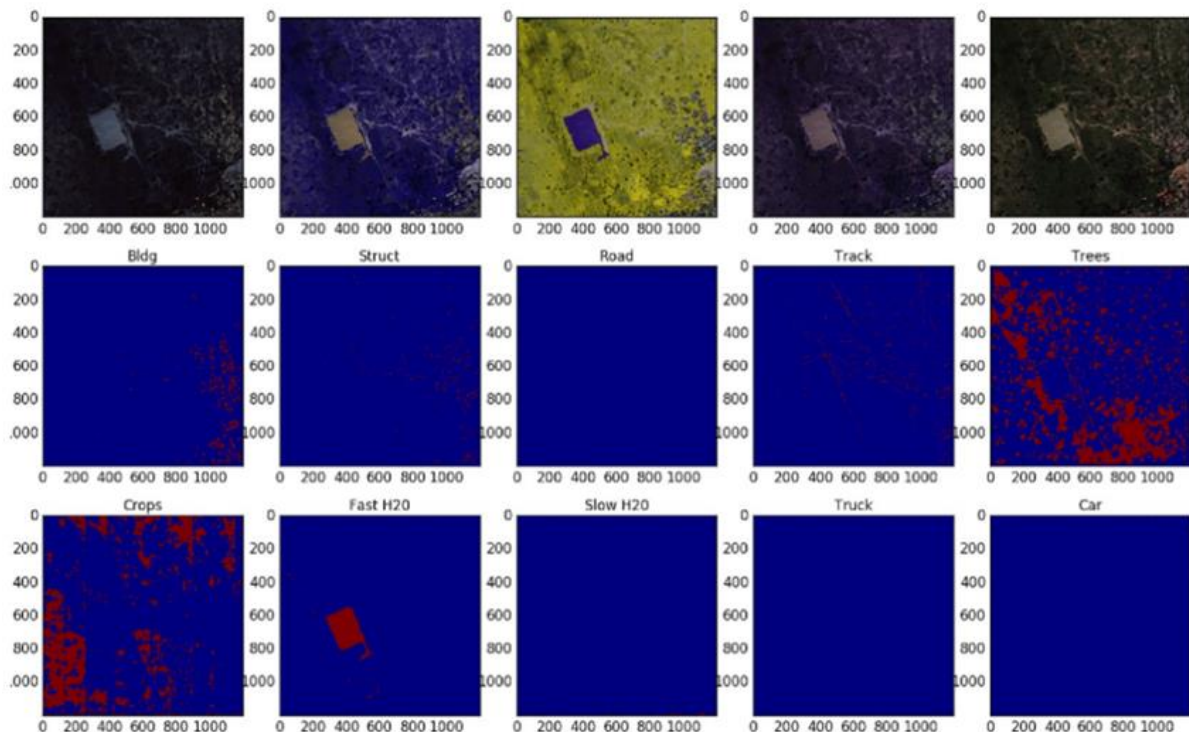


Fig.4. Initial preliminary identification of all categories of images.



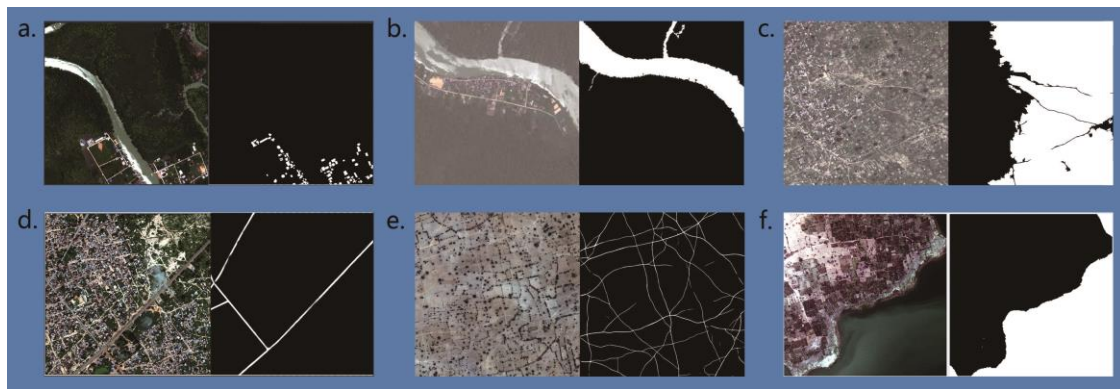


Fig.5. The results of identification

#### 4.EVALUATION OF RESULTS

The data used in this experiment is a set of multi-spectral remote sensing images provided by the German laboratory DSTL on Kaggle. To verify the accuracy of the experimental results, first calculate the Jaccard index of all classes of training data, and then use the network model to evaluate Training data, and then submit the assessment results to kaggle backstage for verification. Table 2 shows the corresponding Jaccard coefficient for each type of feature in the training data.

Category	Training area	Polygonal area	Jaccard coefficient
Housing	6.843670e-05	0.002070227	3.305758e-02
Small building	1.460820e-05	0.002070227	7.056327e-03
Main road	1.674853e-05	0.002070227	8.090188e-03
Other trails	6.217386e-05	0.002070227	3.003239e-02
Trees	2.106242e-04	0.002070227	1.017396e-01
Crops	5.706590e-04	0.002070227	2.756504e-01
River	1.023189e-05	0.002070227	4.942400e-03
Water area	3.488082e-06	0.002070227	1.684879e-03
Small cars	7.719743e-08	0.002070227	3.728935e-05
Large cars	3.160106e-07	0.002070227	1.526454e-04

Table 2 The Jaccard coefficient of training data

In the above table, we can see that the Category 5 trees accounted for 0.1% of the total, the Category 6 crops accounted for 0.27 as a whole, and the other categories accounted for very few ratios. For these classification results, the size of the Jaccard coefficient is equal to the category. The size of the image area occupied, so we only need to obtain the Jaccard coefficient of each category, we can determine the experimental recognition effect. Table 3 shows the results of the experiment given by the kaggle background.

No.	Category	Public Score	Private Score
1	Housing	0.07732	0.06102
2	Small building	0.01832	0.02126
3	Main road	0.07693	0.05522
4	Other trails	0.03765	0.04114
5	Trees	0.05102	0.06371
6	Crops	0.08079	0.08137
7	River	0.09645	0.08972
8	Water area	0.06179	0.04598
9	Small cars	0.03154	0.00747
10	Large cars	0.00311	0.00429
Final recognition accuracy		0.53129	0.46937

Table 3 Experimental identification results

The public score is the result of verification using 19% of the test data, and the private score is the verification result of the remaining 81% of the data. It is not difficult to find that the recognition rates for houses, road trunk lines, crops and rivers are relatively high, and the recognition rates for automobiles, artificial irregular buildings and artificial trails are relatively low, especially for the identification of large-scale automobiles.

#### ACKNOWLEDGEMENTS

This study was supported by Scientific and Technological Development Scheme of Jilin Province (Grant No. 20180101048JC), Major Special Project of China High-resolution Earth Observation System (CHEOS) (04-Y20A35-9001-15/17)

#### REFERENCES

- Castelluccio, Marco, et al. "Land use classification in remote sensing images by convolutional neural networks." arXiv preprint arXiv:1508.00092 (2015).
- Chen, Yushi, Xing Zhao, and Xiuping Jia. "Spectral-spatial classification of hyperspectral data based on deep belief network." IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 8.6 (2015): 2381-2392.
- Cheng, Gong, Peicheng Zhou, and Junwei Han. "Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images." IEEE Transactions on Geoscience and Remote Sensing 54.12 (2016): 7405-7415.
- Heermann, Philip Dale, and Nahid Khazenie. "Classification of multispectral remote sensing data using a back-propagation neural network." IEEE Transactions on Geoscience and Remote Sensing 30.1 (1992): 81-88.
- Hu, Wei, et al. "Deep convolutional neural networks for hyperspectral image classification." Journal of Sensors 2015 (2015).
- Kussul, Nataliia, et al. "Deep learning classification of land cover and crop types using remote sensing data." IEEE Geoscience and Remote Sensing Letters 14.5 (2017): 778-782.
- Makantasis, Konstantinos, et al. "Deep supervised learning for hyperspectral data classification through convolutional neural networks." Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International. IEEE, 2015.

Vakalopoulou, Maria, et al. "Building detection in very high resolution multispectral data with deep learning features." Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International. IEEE, 2015.

Wang, Lizhe, et al. "Spectral-spatial multi-feature-based deep learning for hyperspectral remote sensing image classification." *Soft Computing* 21.1 (2017): 213-221.

Zhang, Liangpei, Lefei Zhang, and Bo Du. "Deep learning for remote sensing data: A technical tutorial on the state of the art." *IEEE Geoscience and Remote Sensing Magazine* 4.2 (2016): 22-40.

Zou, Qin, et al. "Deep learning based feature selection for remote sensing scene classification." *IEEE Geoscience and Remote Sensing Letters* 12.11 (2015): 2321-2325.

Zhang, Liangpei, et al. "Deep learning for remote sensing image understanding." *Journal of Sensors* 2016 (2016).

Zhao, Wenzhi, et al. "On combining multiscale deep learning features for the classification of hyperspectral remote sensing imagery." *International Journal of Remote Sensing* 36.13 (2015): 3368-3379.